AFRL-SR-AR-TR-05-

0030

REPORT DOCUMENTATION PAGE

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing ins... gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this co... information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE Annual	3. DATES COVERED *(From - To)* September 2003 - August 2004

4. TITLE AND SUBTITLE

Real-Time Fault Tolerant Networking Protocols

5a. CONTRACT NUMBER

5b. GRANT NUMBER

F49620-00-1-0327

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Thomas A. Henzinger

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Califronia, Berkeley

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Office of Scientific Research
4015 Wilson Blvd
Mail Room 713
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)

AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution Statement A. Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We made significant progress in the areas of video streaming, wireless protocols, mobile ad-hoc and sensor networks, peer-to-peer systems, fault tolerant algorithms, dependability and timing analysis, and real-time networking software design and maintainance. The following highlights, as well as additional accomplishments, will be described below in more detail.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Thomas A. Henzinger
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU		

Sjogren Jon A Civ AFRL/AFOSR

From: Tom Henzinger [tah@epfl.ch]
Sent: Thursday, July 29, 2004 7:32 PM
To: Jon Sjogren
Cc: Annie Owens; Erin Reiche; Avideh Zakhor; Nancy Lynch; Kishor Trivedi; Mostafa Ammar; Kang Shin
Subject: Real-Time Networking MURI Annual Progress Report

Dear Jon, here is our annual report. Best regards, Tom

Real-Time Fault-Tolerant Networking Protocols
Multi-University Research Initiative
AFOSR Grant F49620-00-1-0327

Annual Progress Report
Covering September 2003 to August 2004

Principal Investigator: Thomas A Henzinger
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

1. Objectives

Researchers at the University of California, Berkeley (Tom Henzinger and Avideh Zakhor), the Massachusetts Institute of Technology (Nancy Lynch), the Georgia Institute of Technology (Mostafa Ammar), the University of Michigan (Kang Shin), and Duke University (Kishor Trivedi) have teamed up to investigate foundational and experimental techniques for enabling real-time, fault-tolerant network protocols.

Our overall research goal is to study networking architectures, services, and algorithms which require innovative quality-of-service and fault-tolerance mechanisms. We have focused on multimedia delivery in traditional client-server architectures, both in the case of the Internet and wireless networks, as well as on peer-to-peer content delivery and on mobile ad-hoc networks.

The unique composition of the team brings new synergies to the problem domain which permit the complete illumination of each newly proposed protocol from all angles, from mathematical modeling and analysis to experimental evaluation, from real-time and QoS aspects to fault-tolerance and reliability aspects. Our approach is to improve newly designed protocols through feedback from timing and fault analysis, and to develop new analysis techniques driven by new protocol designs.

2. Status of Effort

We made significant progress in the areas of video streaming, wireless protocols, mobile ad-hoc and sensor networks, peer-to-peer systems, fault-tolerant algorithms, dependability and timing analysis, and real-time networking software design and maintainance. The following highlights, as well as additional accomplishments, will be described below in more detail.

In the area of video streaming, we developed a new rate control strategy for wireless streaming, new routing protocols, and a a new technique to determine the playout buffer size required to achieve smooth video playout when the video is streamed over TCP.

In the area of wireless protocols, we continued our exploration into the message-ferrying data routing paradigm in challenged, delay tolerant wireless networks. We developed a distributed airtime-based admission control protocol and a weighted-fair enhancement to the IEEE 802.11.

For ad-hoc networks, we defined new programming abstractions, developed algorithms for message flooding, and a tracking algorithm for sensor networks.

For peer-to-peer systems, we developed and evaluated schemes for reputation tracking, and we built a content protection architecture for decentralized peer-to-peer systems.

In the area of fault-tolerant algorithms, we developed consensus algorithms for optimal resilience with Byzantine shared memory, and wait-free storage from Byzantine components.

In the area of dependability and timing analysis, we used Markov regenerative process models to study the availability of internet-based services, and we developed a general framework for survivability quantification.

In the area of networking software design and maintainance, we developed software rejuvenation policies. We also designed virtual machines and type systems that guarantee the real-time behavior of software.

3. Accomplishments

3.1 Video Streaming

We developed rate control strategy for wireless streaming scenario, which results in full utilization of the wireless channel, while not requiring any modifications to the network infrastructure. The approach is based on opening multiple end to end connections and observing delay as a way to determine the optimal number of connections; Verified the approach experimentally on the Verizon Wireless 1XRTT network, showing increased throughput. Currently developing mathematical framework based on Kelly's work to show the scalibiltiy of the approach.

We developed routing protocols for multi-path, unicast, video streaming over wireless adhoc networks, and verified the results via NS simulations. Developed theoretical connectivity requirements for double tree multicast on wireless adhoc networks and showed that they are not more stringent than that of single tree multicast. Proposed serial and parallel schemes for constructing double tree multicast.

We developed a new multicast content distribution mechanism based on k-DAGs, a special Directed Acyclic Graph, where every receiver has at least k parents, thereby increasing fault tolerance and resiliency to parent loss. We verified the results via simulations.

TCP video streaming to memory-constrained devices: TCP is one of the most popular transport protocols for video streaming, even though the rate variability of TCP makes it difficult to provide good video quality. To accommodate the variability, video streaming applications require receiver-side buffering. In current practice, however, there are no guidelines for the provisioning of the receiver buffer, and smooth playout is insured through over-provisioning. In this work, we are interested in memory-constrained devices where it is important to determine the right size for the playout buffer in order to insure a prescribed video quality. To that end, we characterize video streaming over TCP in a systematic and quantitative manner. We first model video streaming system analytically and derive expressions of buffer size requirements based on the model. Our model takes account of three scenarios: 1) when TCP throughput matches video encoding rate, 2) when TCP throughput is smaller than the encoding rate, and 3) when TCP throughput is limited by the maximum window size. Experimental results of both simulations and the Internet experiments validate the model and demonstrate that the minimum buffering delay achieves desired video quality.

Scalable Live Video Streaming to cooperative clients using time shifting and video patching: We considered the problem of how to enable the streaming of live video content from a single server to a large number of clients. One recently proposed approach relies on the cooperation of the video clients in forming an application layer multicast tree over which the video is propagated. Video continuity is maintained as client departures disrupt the multicast tree, using multiple description coded (MDC) streams multicast over several application layer trees. While this maintains continuity, it can cause video quality fluctuation as clients depart and trees are reconstructed around them. We developed a scheme using the transmission of a single-description coded video over an application layer multicast tree formed by cooperative clients. Video continuity is maintained in spite of tree disruption caused by departing clients using a combination of two techniques: 1) providing time-shifted streams at the server and allowing clients that suffer service disconnection to join a video channel of the time-shifted stream, and 2) using video patching to allow a client to catch up with the progress of a video program. Simulation experiments demonstrate that our design can achieve uninterrupted service, while not compromising the video quality, at moderate cost.

Optimal Quality Adaptation for Scalable Encoded Video: The dynamic behavior of the Internet's transmission resources makes it difficult to provide perceptually good quality streaming video. Scalable video encoding techniques have been proposed to deal with this problem. However, an encoded video generally exhibits significant data rate variability to provide consistent visual quality. We are, therefore, faced with the problem of accommodating the mismatch between the available bandwidth variability and the encoded video variability. In this paper, we investigate quality adaptation algorithms for scalable encoded VBR video over the Internet. Our goal is to develop a quality adaptation scheme that maximizes perceptual video quality by minimizing quality variation while at the same time increasing the usage of available bandwidth. We propose an optimal adaptation algorithm and a real time adaptation algorithm based on whether the network conditions are known a priori. Experimental results show that the real time adaptation as well as the optimal adaptation algorithm provide consistent video quality when used over both TFRC and TCP.

3.2 Wireless Protocols

Message ferrying in delay tolerant networks: The Message Ferrying (MF) scheme has been proposed as a strategy for providing connectivity in sparse partitioned ad hoc networks. A set of nodes called ferries are responsible for carrying messages for all nodes in the networks. A ferry periodically moves around the deployed area along a pre-determined route and relays messages between mobile nodes which otherwise cannot communicate with each other directly. The MF scheme relies on the ferry to provide connectivity and is vulnerable to a single point of failure. Also, mobile nodes might take turns to be a ferry since they normally have limited resources. Therefore, ferry replacement is important for robustness in the MF deployment. We proposed two ferry replacement protocols. The ferry designates its successor in the designation scheme while nodes collaborate and elect one node among themselves as the ferry replacement in the distributed election approach. We evaluated the performance of these two replacement protocols through comprehensive simulation.

We also studied the use of multiple ferries in such networks, which may be necessary to address performance and robustness concerns. We focus on the design of ferry routes. With the possibilities of interaction between ferries, the route design problem is challenging. We presented algorithms to calculate routes such that the traffic demand is met and the data delivery delay is minimized. We evaluated these algorithms under a variety of network conditions via simulations. Our goal is to guide the design of MF systems and understand the tradeoff between the incurred cost of multiple ferries and the improved performance. We showed that the performance scales well with the number of ferries in terms of throughput, delay and resource requirements in both ferries and nodes.

Energy-limited capacity of wireless networks: The performance of large-scale wireless ad hoc networks is often limited by the broadcasting nature of the wireless medium and the inherent node energy constraints. While the impact of the former on network capacity has been studied extensively in the literature, the impact of energy constraints has not received as much attention. We studied the capacity limitations resulting from the energy supplies in wireless nodes. We defined the energy-limited capacity of a wireless network as the maximum amount of data the network can deliver before the nodes run out of energy. This energy-limited capacity is an important parameter in networks where operating lifetime is critical, such as ad hoc networks deployed in hazardous environments and sensor networks. We study two types of sensor networks, either with or without the deployment of base stations that are equipped with unlimited energy to support data forwarding. We derived upper and lower bounds on the energy-limited capacity of these networks. While throughput has been shown to not scale with node density by previous studies, our results show that, depending on the energy consumption characteristics of wireless communication, the energy-limited capacity can scale well in sensor networks. In addition, the deployment of base stations can significantly improve the energy-limited capacity of sensor networks. We also studied ad hoc networks and obtain upper and lower bounds of the energy-limited capacity.

We developed an airtime-based admission control with distributed airtime allocation for QoS support in the multi-rate IEEE802.11e wireless LAN. We developed a unified smooth and fast handoff scheme, based on an enhanced Inter-Access Point Protocol (IAPP), which uniformly facilitates inter- and intra-subnet handoffs. We designed the architecture and protocols for spectral agile wirless networks. We developed an analytical model for performance analysis and implement the spectral agile functionalities in the IEEE 802.11 wireless LANs in ns-2.

We developed smart power-saving mode for IEEE 802.11 WLANs which directs a wireless station to sleep/wake up according to an ``optimal'' sequence, such that the desired delay performance is guaranteed with minimum energy consumption.

We developed Weighted-Fair and Bandwidth-Efficient Enhancement to the IEEE 802.11. The key idea is that the medium access parameters of each wireless station are properly selected to (1) reflect the relative weights among contending stations, so as to achieve the weighted fairness among them; and (2) reflect the number of stations contending for the wireless medium, so as to maximize the aggregate throughput. We proposed an optimal tradeoff approach that (i) constructs a hybrid routing protocol by combining well-known location-update schemes (i.e., proactive location updates within nodes' local regions and a distributed location service), and (ii) derives its optimal configuration, in terms of location-update thresholds (both distance- and time-based), to minimize the overall routing overhead.

3.3 Ad-hoc Networks

A Message Ferrying Approach for Data Delivery in Sparse Mobile Ad Hoc Networks: Mobile Ad Hoc Networks (MANETs) provide rapidly deployable and self-configuring network capacity required in many critical applications, e.g., battlefields, disaster relief and wide-area sensing. In this paper we study the problem of efficient data delivery in sparse MANETs where network partitions can last for a significant period. Previous approaches rely on the use of either long-range communication, whcih leads to rapid drain of mobile nodes' limited battery, or node mobility which suffers low delivery rate and large delays. In this paper we describe a Message Ferrying (MF) approach to address the problem. MF is a mobility-assisted approach which utilizes a set of special nodes called message ferries (or ferries for short) to provide communication service for nodes in the area. The main idea behind the MF approach is to introduce non-randomness in the movement of nodes and exploit such non-randomness to help deliver data. We study two variations of the MF scheme, depending on whether ferries or nodes initiate proactive movement. The MF design exploits the mobility of ferries or nodes to improve data delivery performance and reduce energy consumption in nodes. We evaluate the performance of MF vi extensive ns simulations which show MF's efficiency in both message delivery and energy usage under a variety of network conditions.

GeoQuorums: We defined a new programming abstraction for mobile networks: a virtual network consisting of mobile client nodes and virtual static nodes, each of which is an atomic object at a fixed geographical location. We showed how it can be implemented in a mobile setting, and how it can be used to implement a highly fault-tolerant atomic read/write memory.

Virtual mobile nodes: We continued our work on abstractions for mobile networks, this time defining Virtual Mobile Nodes, which move in a pre-planned manner. We showed how such nodes can be implemented using real mobile nodes, which move unpredictably; for example, cars moving in one direction on a road can implement a virtual node that moves in the opposite direction. We sketched several applications, for example, a routing strategy.

Message flooding in sensor networks: We described a simple algorithm for flooding messages in a mobile or sensor network. The algorithm tolerates a variety of failures and network changes, and achieves quite low communication cost. The basic idea is to combine two kinds of communication: flood newly-acquired information immediately, while monitoring in the background to detect when neighboring nodes should be brought up to date.

Gradient clock synchronization: We formulated an interesting "gradient property" for mobile/sensor network clock synchronization: that clocks of nearby processors always be closely synchronized. We proved a fundamental result saying that this property is impossible to achieve under standard network assumptions. Current work in progress: Construct practical clock synch algorithms for sensor/mobile networks, that make use of GPS when available as well as hardware clocks, and that have the "best possible" gradient behavior (subject to the impossibility result).

STALK: a tracking algorithm for sensor networks: We developed a new algorithm for tracking a moving object in a sensor network. The algorithm relies on a hierarchical organization of the network, produced by another algorithm (of Demirbas and Arora). It builds and maintains a tracking path for each object, using leader nodes at different levels of the hierarchy. It tolerates concurrency and failures; failures are handled efficiently, within the lowest possible level of the hierarchy.

Energy efficient connected clusters for mobile ad hoc networks: The merit of a clustered decomposition for a mobile ad hoc network depends on the application that is meant to use it. A power control based distributed clustering algorithm is proposed that maintains cluster connectivity under reasonable assumptions. The size and sparsity of the clustering is controlled by two parameters: the minimal separation between the clusterheads, and the maximum angular gap between neighboring clusterheads. The optimal value of the latter is derived; this minimizes the transmission power of the clusterheads while guaranteeing connectivity of the cluster graph. Experimental studies show that the algorithm rapidly stabilizes to a new clustered organization after the network topology changes due to node joins and failures.

3.4 Peer-to-Peer Systems

Peer-to-peer system reputation tracking: The success of incentive techniques to motivate freeriders to contribute resources in Gnutella-like peer-to-peer networks depends on the availability of peer behavior tracking in terms of resource consumption and contribution. Though many reputation systems have been proposed toward the goal of behavior tracking, the overheads incurred in such tracking have received little attention. Consideration of overheads is an important factor in judging the merits of a practical reputation tracking scheme in order to maintain the scalability of the underlying peer-to-peer network. This paper proposes two methods of reputation tracking: strong and weak reputations. These methods differ in the trade-offs between reliability of reputation tracking and the overheads incurred. We formally specify and verify strong reputations and note that while the scheme yields highly reliable reputation tracking, the reliability and overhead trade-offs in weak reputations may present a more viable alternative for certain applications.

Service Differentiation in Peer-to-Peer Networks Utilizing Reputations: As the population of P2P networks increases, service differentiation issues become very important in distinguishing cooperating peers from free-loaders. The basis for service differentiation could either be economic or the fact that the peers differ from each other in the type of services and resources they contribute to the system. Taking the latter approach, this paper makes three contributions: 1) it defines parameters that are suitable for service differentiation in P2P networks, 2) it proposes SDP, a protocol to accomplish service differentiation, and 3) it identifies a set of features that are necessary in a reputation system that measures the contributions of individual peers in the system.

Video streaming in peer-to-peer systems: We developed a novel architecture for streaming video in a dynamic peer-to-peer system using simple, and standards-based single-description video coding.

CITADEL: A Content protection Architecture for Decentralized Peer-to-Peer Systems. There is an increased interest, by content creators and owners, in content protection systems that provide the ability to control or restrict the content that can be shared on peer-to-peer file sharing systems. Some content protection systems have been proposed for centralized peer-to-peer systems (such as Napster) where a central authority controls all indexing and querying. These systems cannot be applied to decentralized peer-to-peer systems since they rely on a central server. Also, such systems limit the ability of end-users to effectively share content and can make the peer-to-peer distribution model resemble a client-server model in many respects. We proposed CITADEL, a novel content protection architecture designed to operate in decentralized peer-to-peer systems (such as Gnutella). CITADEL enforces a range of protection policies while maintaining an open peer-to-peer distribution model. CITADEL builds a protected file sharing environment over a normal peer-to-peer network using secured content objects and file sharing software enhanced to perform protection operations. A flexible content importation system that is part of CITADEL allows all users to insert new content as well as additional copies of protected content.

A File-Centric Model for Peer-to-Peer File-Sharing Systems: Peer-to-peer systems have quickly become a popular way for file sharing and distribution. We have focused on the subsystem consisting of peers and their actions relative to a specific file and develop a simple theoretical file-centric model for the subsystem. We began with a detailed model that tracks the complete system state. To deal with the large system state space, we investigated a decomposed model, which not only greatly reduces the complexity of solving the system, but also provides a flexible framework for modeling multiple classes of peers and new system features. Using the model, we can study performance measures of a system, such as throughput, success probability of a file search, and the number of file replicas in the system. Our model can also be used to understand the impact of user behavior and new system features. As examples, we investigated the effect of freeloaders, hold-enabled applications and decoys.

3.5 Fault-Tolerant Algorithms

Optimal resilience with byzantine shared memory: We developed Byzantine Disk Paxos, an asynchronous shared-memory consensus protocol that uses a collection of n>3t disks, t of which may fail by becoming non-responsive or arbitrarily corrupted. We gave two constructions of this protocol, each one based on a different building block. One building block is a shared wait-free safe register. The second building block is a regular register that satisfies a new weak termination (liveness) condition, called finite-write terimnation (FW-termination). We constructed each of these reliable registers from n>3t base registers, t of which can be non-responsive or Byzantine. This improves the failure resilience of all the previous wait-free constructions in this model.

Optimal resilience wait-free storage from byzantine components: We described the results of our on-going investigation into the possibility and cost of building a survivable store. We considered optimal resilience systems comprising of 3t+1 base storage units, t of which may fail by becoming non-responsive or arbitrarily corrupted. Our contribution includes both algorithms and lower bounds in this model.

Active Disk Paxos with infinitely many processes: We presented an improvement to the Disk Paxos protocol by Gafni and Lamport that utilizes extended functionality and flexibility provided by Active Disks and supports unmediated concurrent data access by an unlimited number of processes. The solution facilitates coordination by an infinite number of clients using finite shared memory. It is based on a collection of read-modify-write objects with faults, that emulate a new, reliable shared memory abstraction called a ranked register.

Light-weight leases for storage-centric coordination: We proposed light-weight lease primitives to leverage fault-tolerant coordination among clients accessing a shared storage infrastructure (such as network attached disks or storage servers). In our approach, leases are implemented from the very shared data that they protect. That is, there is no global lease manager, there is a lease per data item (e.g., a file, a directory, a disk partition, etc.) or a collection thereof. Our lease primitives are useful for facilitating exclusive access to data in systems satisfying certain timeliness constraints. In addition, they can be utilized as a building block for implementing dependable services resilient to timing failures. In particular, we show a simple lease based solution for fault-tolerant Consensus which is a benchmark distributed coordination problem. Currently, an effort is under way to model and verify the algorithms using the new Timed IOA (TIOA) framework (see below).

Wait-free regular storage from byzantine components: We presented a simple, efficient, and self-contained construction of a wait-free regular register from Byzantine storage components. Our approach combines techniques from the literature on wait-free register constructions with the Byzantine-resilient storage replication methodology. Our construction utilizes a novel building block, called 1-regular register, which can be implemented from Byzantine fault-prone components with the same round complexity as a safe register, and with only a slight increase in storage space.

3.6 Dependability and Timing Analysis

TIOA modeling framework for timed systems: We have assembled a comprehensive monograph on the TIOA model, formulated to be consistent with (a special case of) our recently-developed Hybrid I/O Automata modeling framework. The monograph includes the basic theory, including composition, levels of abstraction, rely-guarantee reasoning, safety vs. liveness, region constructions, etc. Everything is illustrated with simple examples. This is intended to be useful as a general handbook about timed system modeling, for both theoreticians and system developers. It includes the complete theory, as well as suggestions for how to use it to model systems.

Rely-guarantee reasoning for TIOA: We recently extracted a paper from the monograph, this one focusing on rely-guarantee reasoning for timed systems. The main result is a 3-level rely-guarantee technique: to show that A1 || B1 implements A2 || B2, we first weaken A2 and B2 to A3 and B3, respectively. E.g., B3 captures exactly the assumptions about B2 that are needed to show that A1 implements A2. We developed the technique for both safety and liveness properties.

HIOA stability analysis: Sayan Mitra has been continuing his work on analysis methods for systems modeled as Hybrid I/O Automata. This year, he incorporated some stability analysis methods from control theory into his set of methods. For example, working with control theorist Daniel Liberzon, he formulated Liberzon's mixed-mode stability methods within HIOA, and applied them to simple control examples.

PIOA new compositionality results: We have made considerable progress this year on a compositional modeling framework for (asynchronous, nondeterministic) probabilistic systems. First, we have elucidated the problems with earlier definitions for composition and external behavior for PIOAs: those definitions allow the entity that schedules different components so much power that the entire internal branching structure of the PIOAs is exposed. Our new solution to this problem is to restrict the scheduler's power so that its choices can depend only on externally-visible behavior of the components. As an important first step, we have considered a restricted form of PIOA, "switched automata". Switched automata explicitly control their own scheduling, passing control from one to the other via special control actions. We have proved powerful compositionality results for this restricted model. Moreover, we believe that this model forms the right basis for a compositional theory for more general PIOAs---we are working on handling general PIOAs by passing systematically to "switched versions".

Modeling of user perceived Webserver availability: We propose to use

Markov regenerative process (MRGP) models to study the availability of Internet-based services perceived by a Web user, which capture the interactions between the service facility and the user. The necessity of the sophisticated MRGP modeling is evidenced by the comparisons with the corresponding continuous time Markov chain (CTMC) models, which show that the popular convenient CTMC models tend to overestimate user-perceived service unavailabilities by 26% to 125%. We study two different online service scenarios: (1) single-user-single-host and (2) single-user-multiple-host. It is found that user-perceived service unavailability depends not only on the infrastructure's failure-recovery characteristics but also, more importantly, on the user's behavior. Also, for a service provider, to improve users¡¯ satisfaction, inventing a fast recovery mechanism is more effective than striving for a more reliable platform given the platform availability is the same.

Model-Based Evaluation: From Dependability to Security: The development of techniques for quantitative, model-based, evaluation of computer system dependability has a long and rich history. A wide array of stochastic techniques are now available, ranging from combinatorial methods, which are useful for quick, rough-cut, analyses, to state-based methods, such as Markov reward models, and detailed, discrete-event, simulation. The use of quantitative techniques for security evaluation is much less common, and has typically taken the form of formal analysis of small parts of an overall design, or experimental ¡°red team¡±-based approaches. Alone, neither of these approaches is fully satisfactory, and we argue that there is much be gained through the development of a sound model-based methodology for quantifying the security one can expect from a particular design. In this work, we survey existing model-based techniques for evaluating system dependability, and summarize how they are now being extended to evaluate system security. We find that many techniques from dependability evaluation can be applied in the security domain, but that significant challenges remain, largely due to fundamental differences between the accidental nature of the faults commonly assumed in dependability evaluation, and the intentional, human, nature of cyber attackers.

Mean Time To Failure Computation for Non-Coherent Systems: A new algorithm based on Binary Decision Diagram (BDD) is proposed for the computation of steady-state mean time to failure (MTTF) of non-coherent repairable systems. We show the efficiency of our algorithm by applying it to some example fault trees, real-life applications, and large fault tree benchmarks.

A Proactive Approach Towards Always-On Availability in Broadband Cable Networks: We propose a high availability design of a CMTS (Cable Modem Termination System) cluster system based on the software rejuvenation technique. This proactive system maintenance technique is aimed to reduce system outages and the associated downtime cost due to the ``software aging" phenomenon. Different rejuvenation policies are studied from the perspectives of design, implementation, and availability assessment. To evaluate these policies, stochastic reward net models are developed and solved by SPNP (Stochastic Petri Net Package). Numerical results show that the deployment of software rejuvenation in the system leads to significant improvement in capacity-oriented availability and reduction in downtime cost. The optimization of the rejuvenation interval in the time-based approach and the effect of the prediction coverage in the measurement-based approach are also studied in this paper.

A general Framework of Survivability Quantification: We propose a general survivability quantification framework which is applicable to a wide range of system architectures, applications, failure/recovery behaviors, and desired metrics. We show how this framework can be used to derive survivability measures based on different definitions and extend it to other measures not covered by current definitions which can provide helpful information for better understanding of system steady state and transient behaviors under failures/attacks. An illustrative example of a telecommunications switching system is. given for the ease of discussion. Markov models are developed and solved to depict various aspects of system survivability.

Survivability Analysis of Telephone Service: The telecommunications industry has achieved high reliability and availability for telephone service over decades of development. However, the current design does not aim at providing service survivability when a local switching office fails due to catastrophic damage. In this paper, several survivable architectures for telephone subscriber network are proposed based on common survivability principles. In order to quantitatively assess the effectiveness of design alternatives, a set of analytical models are developed to derive various survivability measures. Numerical results are provided to show how a comprehensive understanding of the system behavior after failure can be achieved through different survivability aspects.

Model Checking Discounted Temporal Properties: Temporal logic is two-valued: a property is either true or false. When applied to the analysis of stochastic systems, or systems with imprecise formal models, temporal logic is therefore fragile: even small changes in the model can lead to opposite truth values for a specification. We present a generalization of the branching-time logic CTL which achieves robustness with respect to model perturbations by giving a quantitative interpretation to predicates and logical operators, and by discounting the importance of events according to how late they occur. In every state, the value of a formula is a real number in the interval [0,1], where 1 corresponds to truth and 0 to falsehood. The boolean operators and and or are replaced by min and max, the path quantifiers E and A determine sup and inf over all paths from a given state, and the temporal operators F and G specify sup and inf over a given path; a new operator averages all values along a path. Furthermore, all path operators are discounted by a parameter that can be chosen to give more weight to states that are closer to the beginning of the path. We interpret the resulting logic DCTL over transition systems, Markov chains, and Markov decision processes. We present two semantics for DCTL: a path semantics, inspired by the standard interpretation of state and path formulas in CTL, and a fixpoint semantics, inspired by the mu-calculus evaluation of CTL formulas. We show that, while these semantics coincide for CTL, they differ for DCTL, and we provide model-checking algorithms for both semantics.

The Element of Surprise in Timed Games: We consider concurrent two-person games played in real time, in which the players decide both which action to play, and when to play it. Such timed games differ from untimed games in two essential ways. First, players can take each other by surprise, because actions are played with delays that cannot be anticipated by the opponent. Second, a player should not be able to win the game by preventing time from diverging. We present a model of timed games that preserves the element of surprise and accounts for time divergence in a way that treats both players symmetrically and applies to all omega-regular winning conditions. We prove that the ability to take each other by surprise adds extra power to the players. For the case that the games are specified in the style of timed automata, we provide symbolic algorithms for their solution with respect to all omega-regular winning conditions. We also show that for these timed games, memory strategies are more powerful than memoryless strategies already in the case of reachability objectives.

3.7 Networking Software Design and Maintainance

Software rejuvenation policies for cluster systems under varying workload: We have analyzed two software rejuvenation policies of cluster server systems under varying workload, called fixed rejuvenation and delayed rejuvenation. In order to achieve a higher average throughput, we proposed the delayed rejuvenation policy, which postpones the rejuvenation of individual nodes until off-peak hours. Analytic models using the well known paradigm of Markov chains are used. Since the size of the Markov model is nontrivial, automated specification generation, and the solution via stochastic Petri nets is utilized. Deterministic time to trigger rejuvenation is approximated by a 20-stage Erlangian distribution. Based on the numerical solutions of the models, we found that under the given context, although the fixed rejuvenation occasionally yields a higher throughput, the delayed rejuvenation policy seems to outperform fixed rejuvenation policy by up to 11%. We also compared the steady-state system availabilities of these two rejuvenation policies.

Analysis of a two-level software rejuvenation policy: A two-level rejuvenation policy for software systems with degradation process is studied. Both full restarts and partial restarts are considered in this rejuvenation strategy. A semi-Markov process model is constructed, and based on its closed-form solution we obtain the system availability as a bivariate function. Then, the rejuvenation policy is analyzed to maximize the system availability. Several different scenarios of software rejuvenation strategy are demonstrated by numerical examples.

A Measurement-Based Model for Software Rejuvenation: Recently, the phenomenon of software aging, one in which the state of the software system degrades with time, has been reported. This phenomenon, which may eventually lead to system performance degradation and/or crash/hang failure, is the result of exhaustion of operating system resources, data corruption and numerical error accumulation. To counteract software aging, a technique called software rejuvenation has been proposed, which essentially involves occasionally terminating an application or a system, cleaning its internal state and/or its environment and restarting it. We first include faults attributed to software aging in the framework of traditional software fault classification (deterministic and transient), and study the treatment and recovery strategies for each of the fault classes. This helps us understand the nature of software faults and their impact on system availability and performance and aid in choosing the best possible recovery strategy when a fault is triggered. We then propose a measurement-based model to estimate the rate of exhaustion of operating system resources both as functions of time and of the system workload state. A semi-Markov reward model is constructed based on workload and resource usage data collected from the UNIX operating system. We first identify diifferent workload states using statistical cluster analysis and build a state-space model. Corresponding to each resource, a reward function is then defined for the model based on the rate of resource depletion in diifferent states. The model is then solved to obtain estimated times to exhaustion for the resources. The results of the semi-Markov reward model are then fed into a higherlevel availability model that accounts for failure followed by reactive recovery as well as proactive recovery. This model is then used to derive optimal rejuvenation schedules that maximize availability or minimize downtime cost.

A Workload-based Analysis of Software Aging and Rejuvenation: We present a hierarchical model for the analysis of proactive fault management in the presence of system resource leaks. At the low level of the model hierarchy is a degradation model, in which we use a non-homogeneous Markov chain to establish an explicit connection between resource leaks and the failure rate. With the degradation model we prove that the failure rate is asymptotically constant in the absence of resource leaks and it is increasing as leaks occur and accumulate, which confirms the resource leaks as a aging source. The proactive fault management (PFM) is modeled at the higher level as a semi-Markov process. The PFM model takes as input the degradation analysis from the low-level model and allows us to determine an optimal schedule with respect to various system measures.

Schedule Carrying Code: We introduce the paradigm of schedule-carrying code (SCC). A hard real-time program can be executed on a given platform only if there exists a feasible schedule for the real-time tasks of the program. Traditionally, a scheduler determines the existence of a feasible schedule according to some scheduling strategy. With SCC, a compiler proves the existence of a feasible schedule by generating executable code that is attached to the program and represents its schedule. An SCC executable is a real-time program that carries its schedule as code, which is produced once and can be revalidated and executed with each use. We evaluate SCC both in theory and practice. In theory, we give two scenarios, of nonpreemptive and distributed scheduling for Giotto programs, where the generation of a feasible schedule is hard, while the validation of scheduling instructions that are attached to the programs is easy. In practice, we implement SCC and show that explicit scheduling instructions can reduce the scheduling overhead up to 35% and can provide an efficient, flexible, and verifiable means for compiling Giotto programs on complex architectures, such as the TTA.

A Typed Assembly Language for Real-Time Programs: We present a type system for E code, which is an assembly language that manages the release, interaction, and termination of real-time tasks. E code specifies a deadline for each task, and the type system ensures that the deadlines are path-insensitive. We show that typed E programs allow, for given worst-case execution times of tasks, a simple schedulability analysis. Moreover, the real-time programming language Giotto can be compiled into typed E code. This shows that typed E code identifies an easily schedulable yet expressive class of real-time programs. We have extended the Giotto compiler to generate typed E code, and enabled the run-time system for E code to perform a type and schedulability check before executing the code.

Event-driven Programming with Logical Exceution Times: We present a new high-level programming language, called xGiotto, for programming applications with hard real-time constraints. Like its predecessor, xGiotto is based on the LET (logical execution time) assumption: the programmer specifies when the outputs of a task become available, and the compiler checks if the specification can be implemented on a given platform. However, while the predecessor language Giotto was purely time-triggered, xGiotto accommodates also asynchronous events. Indeed, through a mechanism called event scoping, events are the main structuring principle of the new language. The xGiotto compiler and run-time system implement event scoping through a tree-based event filter. The compiler also checks programs for determinism (absence of race conditions) and time safety (schedulability).

4. Supported Personnel

4.1 Faculty

Mostafa Ammar, Georgia Institute of Technology.
Thomas Henzinger, University of California, Berkeley.
Nancy Lynch, Massachussetts Institute of Technology.
Kang Shin, University of Michigan.
Kishor Trivedi, Duke University.
Avideh Zakhor, University of California, Berkeley.

4.2 PhD Students

Duke University: Kalyan Vaidyanathan, Dongyan Chen.

Georgia Institute of Technology: Minaxi Gupta, Meng Guo, Paul Judge, Taehyun Kim, Wenrui Zhao, Li Zou, Jeonghwa Yang.

Massachussetts Institute of Technology: Seth Gilbert, Tina Nolte, Omar Bakr, Rui Fan, Carl Livadas, Joshua Tauber, Sayan Mitra, Ed Solovey, Christine Robson.

University of California, Berkeley: Brian Godfrey, Thinh Nyguen, Minghua Chen, Wei Wei, Puneet Mehra, Vinayak Prabhu, S Matic, B Horowitz, R Majumdar, A Ghosal, Matulya Bansal.

University of Michigan: Daji Qiao, Chun-Ting Chou, Kyu-Han Kim.

4.3 Postdocs

S Dharmaraja, Duke University.
Y Hong, Duke University.
Paul Judge, Georgia Institute of Technology.
Sang Kang, University of California, Berkeley.
Dilsun Kaynar, Massachussetts Institute of Technology.
Idit Keidar, Massachussetts Institute of Technology.
Gregory Chockler, Massachussetts Institute of Technology.
Christoph Kirsch, University of California, Berkeley.
Xiaomin Ma, Duke University.
Marco Sanvido, University of California, Berkeley.
David Harrison, University of California, Berkeley.

5. Publications

5.1 Journal Publications

[1] P. Schmid-Saugeon and A. Zakhor, "Dictionary Design for Matching Pursuit and Application to Motion Compensated Video Coding" in IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, no. 6, June 2004, pp. 880 - 886.

[2] T. Nguyen and A. Zakhor, "Multiple Sender Distributed Video Streaming" in IEEE Transactions on Multimedia, Vol. 6, No. 2, April 2004, pp. 315 - 326.

[3] C. De Vleeschouwer and A. Zakhor, "In loop atom modulus quantization for matching pursuit and its applications to video coding" in IEEE Transactions on Image Processing, Vol. 12, No. 10, October 2003, pp. 1226 - 1242.

[4] P. Mehra, C. De Vleeschouwer, and Avideh Zakhor, "Receiver-driven bandwidth sharing for TCP and its application to video streaming", accepted for publication in IEEE Transactions on Multimedia, October 2003.

[5] S. Cheung and A. Zakhor, "Towards building a similarity video search engine for the world wide web", accepted for publication in IEEE Transactions on Multimedia, December 2003.

[6] S. Kang and A. Zakhor, "Effective Bandwidth Based Scheduling for Streaming Media", accepted for publication in IEEE Transactions on Multimedia, April 2004.

[7] M. Bansal, and A. Zakhor, "Resilient Overlay Multimedia Multicast", submitted to IEEE Transactions on Multimedia Networking.

[8] W. Xie, Y. Hong, and K. Trivedi, "Analysis of a two-level software rejuvenation policy", International Journal on Reliability Engineering

and System Safety, 2004, to appear.

[9] D. M. Nicol, W. H. Sanders, and K. S. Trivedi "Model-Based Evaluation: From Dependability to Security", invited paper in the inaugural issue of IEEE Transaction on Dependable and Secure Computing, to appear.

[11] D. Wang and K. Trivedi, "Mean Time To Failure Computation for Non-Coherent Systems", submitted to IEEE Transaction on Reliability.

[12] Y. Bao, X. Sun, and K. S. Trivedi, "A Workload-based Analysis of Software Aging and Rejuvenation", submitted to IEEE Transaction on Reliability.

[13] K. Vaidyanathan and K. S. Trivedi, "A Measurement-Based Model for Software Rejuvenation", submitted to IEEE Transaction on Dependable and Secure Computing.

[14] Taejoon Park and Kang G. Shin, ``On maximizing bandwidth-efficiency of location-based routing in ad hoc networks,'' IEEE/ACM Trans. on Networking (in press).

[15] Chansoo Yu, Kang G. Shin, and Ben Lee, ``Power-stepped protocol: A novel way of enhancing spatial utilization in a clustered mobile ad hoc network,'' IEEE Journal of Special Areas of Communications: Special Issue on Quality of Service Delivery in Variable topology Networks (in press).

[16] Daji Qiao and Kang G. Shin, ``Weighted-fair and bandwidth-efficient enhancement to the IEEE 802.11 DCF,'' accepted for publication in IEEE Trans. on Mobile Computing subject to minor revision.

[17] Chun-Ting Chou and Kang G. Shin, ``Analysis of adaptive bandwidth allocation in wireless networks with multi-level degradable quality of service,'' IEEE Trans. on Mobile Computing, vol. 3, no. 1, pp. 5--17, Jan.-Mar. 2004.

[18] Taehyun Kim, Mostafa Ammar, "Optimal Quality Adaptation for Scalable Encoded Video," to appear in IEEE Journal on Selected Areas in Communication.

[19] Gregory Chockler and Dahlia Malkhi. Active Disk Paxos with infinitely many processes. To appear in Distributed Computing, 2004.

[20] G. Chockler and D. Malkhi. Light-Weight Leases for Storage-Centric Coordination. Technical Report MIT-LCS-TR-934, MIT Laboratory for Computer Science, January, 2004. Revised April 2004. Invited for submission to International Journal of Parallel Programming.

[21] Toh Ne Win, Michael Ernst, Stephen Garland, Dilsun Kirli, and Nancy Lynch. Using simulated execution in verifying distributed algorithms. International Journal on Software Tools for Technology Transfer (STTT), 4:(1-10), 2003.

5.2 Reviewed Conference Proceedings

[22] M. Chen and A. Zakhor, "Transmission Protocols for Streaming Video over Wireless" in International Conference on Image Processing 2004, Singapore 2004.

[23] W. Wei and A. Zakhor, "Robust Multipath Source Routing Protocol (RMPSR) for Video Communication over Wireless Ad Hoc Networks" in International Conference on Multimedia and Expo 2004, Taipei, Taiwan, June 2004.

[24] W. Wei and A. Zakhor, "Connectivity for Multiple Multicast Trees in Ad Hoc Networks" in International Workshop on Wireless Ad-hoc Networks, Oulu, Finland, June 2004.

[25] M. Chen and A. Zakhor, "Rate Control for Streaming Video over Wireless" in INFOCOM 2004, Hong Kong, April 2004.

[26] S. H. Kang and A. Zakhor, "Effective Bandwidth Based Scheduling for Streaming Multimedia" in International Conference on Image Processing 2003, Barcelona, Spain, September 2003.

[27] S. Cheung and A. Zakhor, "Fast Similarity Search on Video Signatures" in International Conference on Image Processing 2003, Barcelona, Spain, September 2003.

[28] T. Nguyen and A. Zakhor, "Matching Pursuits Based Multiple Description Video Coding for Lossy Environments" in International Conference on Image Processing 2003, Barcelona, Spain, September 2003.

[29] W. Xie, Y. Hong, and K. Trivedi, "Software rejuvenation policies for cluster systems under varying workload", In IEEE Pacific Rim International Symposium on Dependable Computing (PRDC), March, 2004.

[30] W. Xie, H. Sun, Y. Cao, and K. Trivedi, "Modeling of user perceived Webserver availability", In International Conference Communications (ICC), pages 1796-1800, Anchorage, AK, May 2003.

[31] Y. Liu, Y. Ma, J. J. Han, H. Levendel, and K. S. Trivedi "A Proactive Approach Towards Always-On Availability in Broadband Cable Networks", submitted after revision to Communication Networks, fast abstract in 6th International Workshop on Performability Modeling of Computer and Communication Systems (PMCCS-6), Sept. 2003, Monticello, Illinois, USA.

[32] Y. Liu and K. S. Trivedi, "A general Framework of Survivability Quantification", to appear in the 12th GI/ITG Conference on Measuring, Modelling and Evaluation of Computer and Communication Systems 2004, Dresdon, Germany.

[33] Y. Liu, M. Veena, and K. S. Trivedi, "Survivability Analysis of Telephone Service", to appear in IEEE International. Symposium on Software Engineering (ISSRE'04), Nov. 2004 Saint-Malo, Bretagne, France.

[34] Puneet Sharma, Jack Brassil, Sung-Ju Lee, and Kang G. Shin, ``Intelligent bandwidth aggregation for mobile collaborative communities,'' Proc. Broadband Networks 2004---Broadband Wireless Networking Symposium (in press).

[35] Puneet Sharma, Jack Brassil, Sung-Ju Lee, and Kang G. Shin, ``Distributed channel monitoring for wireless bandwidth aggregation,'' Proc. IFIP-TC6 Networking Conference 2004, Athens, Greece, May 2004.

[36] Chun-Ting Chou, Kang G. Shin, and Sai Shankar, ``Inter Frame Space (IFS) based service differentiation for IEEE 802.11e wireless LANs,'' Proc. 57-th IEEE Vehicle Technology Conference (VTC) 2003-Fall, October 2003.

[37] Daji Qiao, Sunghyun Choi, Amit Jain, and Kang G. Shin, ``MiSer: an optimal low-energy communication strategy for IEEE 802.11a/h,'' Proc. ACM MobiCom'03, pp.161-175, September 16-18, 2003.

[38] Amit Jain, Daji Qiao, and Kang G. Shin, ``RT-WLAN: A soft real-time extension to the ORiNOCO Linux device driver,'' Proc. 14-th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'03), September 7-10, Beijing, China.

[39] Ittai Abraham, Gregory Chockler, Idit Keidar and Dahlia Malkhi. Byzantine Disk Paxos: Optimal Resilience with Byzantine Shared. To appear in the Proceedings of the 23rd Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC 2004) July 25-28, 2004, St. John's, Newfoundland, Canada. ACM, (2004).

[40] Ittai Abraham, Gregory Chockler, Idit Keidar, and Dahlia Malkhi. Wait-Free Regular Storage from Byzantine Components. Submitted for publication, 2004.

[41] Mohammad Taghi Hajiaghayi, Nicole Immorlica, and Vahab S. Mirrokni. Power optimization in fault-tolerant topology control algorithms for wireless multi-hop networks. MOBICOM 2003: Proceedings of the Ninth Annual ACM International Conference on Mobile Computing and Networking, pages 300-312, San Diego, CA, September 2003.

[42] Dilsun K. Kaynar and Nancy A. Lynch. Decomposing Verification of Timed I/O Automata. To appear in FORMATS-FTRTFT, joint conference on Formal Modeling and Analysis of Timed Systems and Formal Techniques in Real-time and Fault Tolerant Systems, Grenoble, France, September 22-24, 2004.

[43] Dilsun K. Kaynar, Nancy Lynch, Roberto Segala, and Frits Vaandrager. Timed I/O Automata: A Mathematical Framework for Modeling and Analyzing Real-Time Systems. RTSS 2003: The 24th IEEE International Real-Time Systems Symposium, Cancun, Mexico, December, 2003.

[44] Roger Khazan and Nancy Lynch. An Algorithm for an Intermittently Atomic Data Service Based on Group Communication. Proceedings of the International Workshop on Large-Scale Group Communication, Florence, Italy, pages 25-30, October 2003.

[45] Carl Livadas and Idit Keidar. Caching-Enhanced Scalable Reliable Multicast. To appear in the International Conference on Dependable Systems and Networks (DSN), June-July 2004.

[46] Nancy Lynch, Roberto Segala, and Frits Vaandrager. Compositionality for Probabilistic Automata. In Roberto Amadio and Denis Lugiez, editors, CONCUR 2003 - Concurrency Theory (14th International Conference on Concurrency Theory, Marseille, France, September, 2003), volume 2761 of Lecture Notes in Computer Science, pages 208-221, Springer-Verlag, 2003. Also, long version to appear as Technical Report MIT-LCS-TR-907, MIT Laboratory for Computer Science, Cambridge, MA 02139.

[46] Sayan Mitra and Myla Archer. Reusable PVS Proof Strategies for Proving Abstraction Properties of I/O Automata. In STRATEGIES 2004, 5th International Workshop on strategies in automated deduction, Cork Ireland, July 2004.

[47] Sayan Mitra and Jesse Rabek. Energy Efficient Connected Clusters for Mobile Ad Hoc Networks. MED-HOC-NET 2004, Third Annual Mediterranean Ad Hoc Networking Workshop, Bodrum, Turkey, June 2004.

[48] Sayan Mitra and Daniel Liberzon. Stability of Hybrid Automata with Average Dwell Time: An Invariant Approach. Submitted for publication.

[49] Sayan Mitra and Myla Archer. Developing Strategies for Specialized Theorem Proving about Untimed, Timed, and Hybrid I/O Automata. In STRATA 2003, Design and Application of Strategies/Tactics in Higher Order Logics, Rome, Italy, September, 2003.

[50] Athicha Muthitacharoen, Seth Gilbert, and Robert Morris. Atomic

Mutable Data in a Distributed Hash Table with Etna. Submitted for publication.

[51] Joshua A. Tauber and Nancy A. Lynch and Michael J. Tsai. Compiling IOA without Global Synchronization. To appear in Proceedings of the 3rd IEEE International Symposium on Network Computing and Applications (IEEE NCA04), Cambridge, MA, August 2004.

[52] Michael A. Bender, Jeremy T. Fineman, Seth Gilbert, Charles E. Leiserson. On-the-Fly Maintenance of Series-Parallel Relationships in Fork-Join Multithreaded Programs. To appear in the Sixteenth ACM Symposium on Parallelism in Algorithms and Architectures, Barcelona, Spain, June 27-30, 2004.

[53] Jacob Beal and Seth Gilbert. RamboNodes for the Metropolitan Ad Hoc Network. To appear in Proceedings of the Workshop on Dependability in Wireless Ad Hoc Networks and Sensor Networks, part of the International Conference on Dependable Systems and Networks, Florence, Italy, June-July, 2004. Also, AI Memo: AIM-2003-027.

[54] Ling Cheung, Nancy Lynch, Roberto Segala, and Frits Vaandrager. Switched Probabilistic I/O Automata. Submitted for publication, July 2004.

[55] Gregory Chockler, Idit Keidar and Dahlia Malkhi. Optimal Resilience Wait-Free Storage from Byzantine Components: Inherent Costs and Solutions. To appear in FuDiCo II: S.O.S. Survivability: Obstacles and Solutions. 2nd Bertinoro Workshop on Future Directions in Distributed Computing, 23-25 June 2004 University of Bologna Residential Center Bertinoro (Forli), Italy.

[56] Murat Demirbas, Anish Arora, Tina Nolte, and Nancy Lynch. Brief Announcement: STALK: A Self-Stabilizing Hierarchical Tracking Service for Sensor Networks. To appear in Proceedings of the 23rd Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC 2004) July 25-28, 2004, St. John's, Newfoundland, Canada.

[57] Shlomi Dolev, Seth Gilbert, Nancy Lynch, Alex Shvartsman, and Jennifer Welch. GeoQuorums: Implementing Atomic Memory in Ad Hoc Networks. In Faith Fich, editor, Distributed Computing (DISC 2003: 17th International Symposium on Distributed Computing, Sorrento, Italy, October, 2003), volume 2848 of Lecture Notes in Computer Science, pages 306-320, Springer-Verlag, 2003.

[58] Shlomi Dolev, Seth Gilbert, Nancy Lynch, Elad Schiller, Alex Shvartsman, Jennifer Welch. Brief Announcement: Virtual Mobile Nodes for Mobile Ad Hoc Networks. To appear in Proceedings of the 23rd Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC 2004) July 25-28, 2004, St. John's, Newfoundland, Canada. Also, Technical Report MIT-LCS-TR-937, MIT CSAIL, Cambridge, MA, 2004.

[59] Shlomi Dolev, Seth Gilbert, Nancy A. Lynch, Elad Schiller, Alex A. Shvartsman, and Jennifer L. Welch. Virtual Mobile Nodes for Mobile Ad Hoc Networks. To appear in the 18th International Symposium on Distributed Computing, Trippenhuis, Amsterdam, the Netherlands, October, 2004.

[60] Rui Fan and Nancy Lynch. Gradient Clock Synchronization. To appear in Proceedings of the Twenty-Third Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing, St. John's, Newfoundland, Canada, July 25-58, 2004.

[61] Rui Fan and Nancy Lynch. Efficient Replication of Large Data Objects. Distributed Computing (DISC 2003): 17th International Symposium on Distributed Computing, Sorrento, Italy, October, 2003), volume 2848 of Lecture Notes in Computer Science, pages 75-91,

Springer-Verlag, 2003.

[62] Wenrui Zhao, Mostafa Ammar, Ellen Zegura, "A Message Ferrying Approach for Data Delivery in Sparse Mobile Ad Hoc Networks," Proceedings of ACM Mobihoc 2004, Tokyo Japan, May 2004.

[63] Meng Guo, Mostafa Ammar, Scalable Live Video Streaming to cooperative clients using time shifting and video patching," Proceedings of IEEE INFOCOM 2004, Hong Kong, March 2004.

[64] Li Zou, Mostafa Ammar, "A File-Centric Model for Peer-to-Peer File-Sharing Systems," Proceedings of IEEE ICNP 2003, Atlanta, GA, Nov. 2003.

[65] Minaxi Gupta, Mostafa Ammar, ``Service Differentiation in Peer-to-Peer Networks Utilizing Reputations," Proceedings of the Networked Group Communications (NGC) workshop, October 2003.

[66] Paul Judge, Mostafa Ammar, ``CITADEL: A Content protection Architecture for Decentralized Peer-to-Peer Systems," Proceedings of IEEE GLOBECOM 2003, Decemeber 2003

[67] J. Yang, Y. Chen, M. Ammar "Ferry Replacement Protocols in Sparse MANET Message Ferrying Systems" submitted to the First IEEE International Conference on Sensor and Ad hoc Communications and Networks

[68] W. Zhao, M. Ammar, E. Zegura, "The Energy-Limited Capacity of Wireless Networks" submitted to the First IEEE International Conference on Sensor and Ad hoc Communications and Networks

[69] T. Kim, M. Ammar, "Determining Playout Buffer Requirements for Video Streaming over TCP" to Memory-Constrained Devices Submitted to IEEE INFOCOM 2005.

[70] W. Zhao, M. Ammar, E. Zegura "Controlling the Mobility of Multiple Data Transport Ferries in a Delay-Tolerant Network". Submitted to IEEE INFOCOM 2005.

[71] M. Gupta, M. Ammar, M. Ahamad "Trade-offs Between Reliability and Overheads in Peer-to-Peer Reputation Tracking". Submitted to IEEE INFOCOM 2005.

[72] Arkadeb Ghosal, Thomas A. Henzinger, Christoph M. Kirsch, and Marco A.A. Sanvido. Event-driven Programming with Logical Exceution Times. Proceedings of the Seventh International Workshop on Hybrid Systems: Computation and Control (HSCC), Lecture Notes in Computer Science 2993, Springer-Verlag, 2004, pp. 357-371.

[73] Thomas A. Henzinger and Christoph M. Kirsch. A Typed Assembly Language for Real-Time Programs. Proceedings of the Fourth International Conference on Embedded Software (EMSOFT), ACM Press, 2004.

[74] Luca de Alfaro, Marco Faella, Thomas A. Henzinger, Rupak Majumdar, and Marielle Stoelinga. Model Checking Discounted Temporal Properties. Proceedings of the 10th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS), Lecture Notes in Computer Science 2988, Springer-Verlag, 2004, pp. 77-92.

[75] Thomas A. Henzinger, Christoph M. Kirsch, and Slobodan Matic. Schedule Carrying Code. Proceedings of the Third International Conference on Embedded Software (EMSOFT), Lecture Notes in Computer Science 2855, Springer-Verlag, 2003, pp. 241-256.

[76] Luca de Alfaro, Marco Faella, Thomas A. Henzinger, Rupak

Majumdar, and Marielle Stoelinga. The Element of Surprise in Timed Games. Proceedings of the 14th International Conference on Concurrency Theory (CONCUR), Lecture Notes in Computer Science 2761, Springer-Verlag, 2003, pp. 144-158.

5.3 Theses and Technical Reports

[77] Christine Margaret Robson. TIOA and UPPAAL. Masters of Engineering Thesis, MIT Department of Electrical Engineering and Computer Science, Cambridge, MA, May 2004.

[78] Edward Solovey. Simulation of Composite I/O Automata. Masters of Engineering Thesis, MIT Department of Electrical Engineering and Computer Science, Cambridge, MA, August 2003.

[79] Paul Attie, Rachid Guerraoui, Petr Kouznetsov, Nancy Lynch, and Sergio Rajsbaum. Boosting Distributed Service Resilience is Impossible. Manuscript, 2004.

[80] Seth Gilbert, Nancy Lynch, and Alex Shvartsman. RAMBO II: Implementing atomic memory in dynamic networks, using an aggressive reconfiguration strategy. Technical Report MIT-CSAIL-TR-890, CSAIL, Massachusetts Institute Technology, Cambridge, MA, 2004.

[81] Joshua A. Tauber. Verifiable Compilation of I/O Automata without Global Synchronization, To appear as PhD Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, September, 2004.

[82] Joshua A. Tauber and Stephen J. Garland. Definition and Expansion of Composite Automata in IOA. To appear as MIT CSAIL Technical Report, July 2004.

[83] Mandana Vaziri, Joshua A. Tauber, Michael J. Tsai, and Nancy Lynch. Systematic Removal of Nondeterminism for Code Generation in I/O Automata. To appear as MIT CSAIL Technical Report, July 2004.

[84] Seth Gilbert. RAMBO II: Rapidly Reconfigurable Atomic Memory for Dynamic Networks. Masters Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, August 2003.

[85] Dilsun Kaynar, Nancy Lynch, Sayan Mitra, Christine Robson. Design for TIOA Modeling Language. Manuscript, February 2004.

[86] Dilsun K. Kaynar, Nancy Lynch, Roberto Segala, and Frits Vaandrager. The Theory of Timed I/O Automata. Technical Report MIT-LCS-TR-917a, MIT Laboratory for Computer Science, Cambridge, MA, April, 2004.

[87] Carolos Livadas and Nancy A. Lynch. A Reliable Broadcast Scheme for Sensor Networks. Technical Report MIT-LCS-TR-915, MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, August 2003.

[88] Nancy Lynch and Ion Stoica. MultiChord: A resilient namespace management algorithm. To appear as Technical Memo MIT-LCS-TR-936, CSAIL, Massachusetts Institute of Technology, Cambridge, MA 2004.

[89] Shlomi Dolev, Seth Gilbert, Nancy A. Lynch, Alex A. Shvartsman, and ennifer L. Welch. GeoQuorums: Implementing Atomic Memory in Ad Hoc Networks. Selected for special edition of Distributed Computing, (edited by Faith Fich), related to the DISC03 conference, 2004. Also, Technical Report MIT-LCS-TR-900a, CSAIL, Massachusetts Institute of Technology, Cambridge, MA, 2004.

[90] Stephen Garland, Nancy Lynch, Joshua Tauber, and Mandana Vaziri. IOA User Guide and Reference Manual. To appear as Technical Report, MIT CSAIL, July 2004.

[91] Benjamin Horowitz, Giotto: A Time-triggered Language for Embedded Programming. PhD thesis, UC Berkeley, 2003.

5.4 Software Releases

Virtual Embedded and Scheduling Machines, Giotto compiler, and corresponding run-time systems for several target architectures and network simulators. University of California, Berkeley.

IOA code generation: This year our work on code generation for distributed systems from IOA models is coming to fruition. We can now generate runnable code (Java interacting with MPI) for our LAN, automatically and directly from IOA models for distributed algorithms. The algorithm models can be proved correct and analyzed for performance using a range of techniques, including hand analysis, interactive theorem-proving, and model-checking. Thus, we essentially have a method of generating verified code.

TIOA: UPPAAL connection. We are currently producing an extension for the IOA language that incorporates time-passage, using algebraic and differential equations and inequalities to describe "trajectories" of state evolution over time. A sub-language of the general TIOA language has been translated to the UPPAAL source language to enable simulation and model-checking. We are currently working on extending our own simulator and building a new theorem-prover interface (this time to PVS).

Reusable PVS Proof Strategies for Proving Abstraction Properties of I/O Automata. We describe an abstraction specification technique and associated abstraction proof strategies we have developed for timed I/O automata. The new strategies can be used together with existing strategies in the TAME (Timed Automata Modeling Environment) interface to PVS; thus, our new templates and strategies provide an extension to TAME for proofs of abstraction. We illustrate how the extended set of TAME templates and strategies can be used to prove example I/O automata abstraction properties taken from the literature.

6. Interactions and Transitions

6.1 Invited Talks

Lynch. Input/Output Automata: Basic, Timed, Hybrid, Probabilistic. Invited talk at Concur 2003 Sept., 2003

Lynch. Using I/O Automata to Model and Analyze Security Protocols Invited talk at DIMACS workshop on Foundations of Security Protocols June, 2004.

Chockler. Optimal Resilience Wait-Free Storage from Byzantine Components: Inherent Costs and Solutions, Invited talk at Summer Research Institute, The Swiss Federal Institute of Technology Lausanne (EPFL), July 2004, Lausanne, Switzerland.

Chockler. Optimal Resilience Wait-Free Storage from Byzantine Components: Inherent Costs and Solutions, SUN Microsystems, Burlington MA, April 2004.

Shin. Invited presentations at University of Florida, Florida Atlantic University, University of Minnesota, Seoul National University, Korea Advanced Institute of Science and Technology, Swedish Royal Institute of Technology.

Zakhor. Invited paper/talk at the International Conference on Image Processing 2004, Singapore 2004.

Zakhor. Invited paper/talk at the International Conference on Image Processing, 2003.

Ammar. Keynote at the NSF-supported International Workshop on Theoretical and Algorithmic Aspects of Sensor, AD-Hoc Wireless and Peer-to-peer Networks, Feb. 2004.

Ammar. Invited talk at the 2003-2004 Distinguished Lecture Series, at the Institute for Computing, Information and Cognitive Systems, The University of British Columbia, Vancouver -- Feb. 2004.

Henzinger. The Fixed Logical Execution Time Assumption, Workshop on Software Engineering for Embedded Systems: From Requirements to Implementation, Chicago, Illinois, September 2003.

Henzinger. Embedded Software: Better Models, Better Code, Jon Postel Distinguished Lecture, University of California, Los Angeles, February 2004.

Henzinger. The Symbolic Approach to Hybrid Systems, Mathematics and Computer Science Colloquium, Santa Clara University, Santa Clara, California, January 2004.

Henzinger. Embedded Software: Better Models, Better Code, keynote address, 2004 International Conference on Applications and Theory of Petri Nets, Bologna, Italy, June 2004.

6.2 Transitions

Rupak Majumdar PhD graduated in August 2003 and has joined UCLA as Assistant Professor. Ben Horowitz PhD graduated in December 2003 and joined Lawrence Livermore Labs. Carl Livadas PhD graduated in summer 2003 and is now a researcher at Bolt, Beranek and Newman, in the networking group. Daji Qiao completed PhD in December 2003, is now assistant professor at Iowa State University. Li Zou, PhD Received, December 2003, Current Position: Senior Software Engineer -- Brion Technologies, Santa Clara, CA. Minaxi Gupta, PhD defended: July 2004, Accepted position as Assistant Professor, Dept. of Computer Science, Indiana University, Bloomington, IN.

Christoph Kirsch, a postdoc with the project, joined the University of Salzburg, Austria, as Professor. Marco Sanvido, another postdoc, joined VMware.

The IOA toolset is readily available. It is being used, for example, by:

* Dr. Nancy Griffeth, formerly of AT&T and now at Lehmann College. She is using it to model, simulate, and study Internet protocols.

* Dr. Chryssis Georgiou, of University of Cyprus. He is using it to generate actual distributed code for a LAN from IOA distributed algorithm descriptions. With the help of MIT undergrad Panayiotis Mavrommatis he has developed and run implementations of leader election, spanning tree, and broadcast algorithms.

* Dr. Frits Vaandrager (Nijmegen) has been using the IOA toolset in his course on Protocol Validation.

The Giotto toolset is being used for real-time programming courses in Kiel, Germany, Salzburg, Austria, and at UC Berkeley.

Kang Shin gave talks and discussed technology transfer at Intel Santa Clara` (California) and Hillsboro (Oregon), HP Labs at Palo Alto (California), IBM T. J. Watson Research Center, Ford Motor Company's Scientific Research Labs, Hyundai Automotive Technology Center (Korea), Samsung Electronics Software Center (Korea), and Philips Research USA (Briar Cliff, NY). We are closely collaborating with HP, Intel, Samsung, and Philips.

7. Patent Disclosures

None.

8. Honors and Awards

Mostafa Ammar was elected ACM FELLOW in December 2003.

Mostafa Ammar was named Regents' Professor in the Collage of Computing at Georgia Tech.

Paul Judge (MURI Fellow who graduated in Dec. 2002) was named to MIT Technology Review's Magazine top 100 young innovators in 2003.

Kang Shin. 2004 Stephen Attwood Award from the College of Engineering at University of Michigan (the highest honor in the college of engineering).

Kang Shin. 2004 Zhu Kezhen International Lectureship Award.

Gilbert et al. DISC 2003 paper on Geoquorums invited for special journal issue.

Rui Fan. Best student paper, PODC 2004, for "Gradient Clock Synchronization". Invited for special journal issue.

Chockler et al. A paper on light-weight leases invited for a special issue of International Journal of Parallel Programming (IJPP).

Ghosal et al. Event-driven Programming with Logical Exceution Times. Invited to special issue of Formal Methods in System Design for best papers of HSCC 2004.

de Alfaro et al. Model Checking Discounted Temporal Properties. Invited to special issue of Theoretical Computer Science for best papers of TACAS 2004.

Henzinger et al. Schedule Carrying Code. Invited to special issue of ACM Trans. Embedded Systems for best papers of EMSOFT 2003.